Efficient Cross Media Retrieval Using Mixed Generative Based Hashing Method

Miss. Mrunmayi Patil
Department of Computer Engineering,
JSPM's Imperial College Of Engineering Research
Wagholi , Pune

Dr.Vinod Wadne
Department of Computer Engineering,
JSPM's Imperial COllege Of Engineering Research
Wagholi, Pune

***Abstract*—Hash methods are proven useful for a variety of tasks and have sparked great attention in recent years. They have proposed several approaches to capture the similarities between textual, visual, and cross-cultural hashing. However, most existing bag of words methods used to represent textual information. Because words with different shapes can they have a similar meaning, semantic text similarities cannot be well elaborated in these methods. To address these challenges in this paper, introduce a new hashing method, which uses continuous representations of proposed words by capturing the semantic textual similarity level and using a deep conviction network (DBN) to build correlation between different modes. In order to demonstrate the effectiveness of the proposed method, three methods commonly used are to be considered background set in this workbook is used. The experimental results show that the proposed method achieves significantly better results in addition, the effectiveness of the proposed method is similar or superior some other hashing methods.**

***Index Terms*—Fisher vector, SCMH, SIFT Descriptor, Word Embedding, Ranking, Mapping.**

I. Introduction

With the fast development of internet and multimedia, information with various form has become enough smooth, simple and easier to access, modify and duplicate. Information with various forms may have semantic correlation for example a microblogs in Facebook often consist of tag, a video in YouTube is always associated with related description or tag as semantic information inherently consist of data with different modality provide an great emerging demand for the applications like cross media retrieval, image annotation and recommendation system. Therefore, the hash similarity methods which calculates or approximate search suggested and received a remarkable attention in last few years.

The core problem of hash learning is how to formulate underlay co-relation between multiple modality and retain / protect the similarity relation in each respective modalities. Generally hashing method divided into 2 categories: matrix decomposition method and vector based method. Matrix de-composition based hashing method search low dimensional spaces to construct data and quantify the reconstruction co-efficient to obtain binary codes. Such kind of methods avoid graph construction and Eigen decomposition. The drawback with such methods, causes large quantization errors which detonate such performance for large code length and design multi-modal hashing model SCMH which focuses on Image and Text type of data with binary representation Hashing. This method processed text data using Skip gram model and image data using SIFT Descriptor. After it generates hash code using Deep Neural network by avoiding duplicates.

II. Literature Survey

Literature survey is the most important step in any kind of research. Before start developing we need to study the previous papers of our domain which we are working and on the basis of study we can predict or generate the drawback and start working with the reference of previous papers.
In this section, we briefly review the related work on Tag Search and Image Search and their different techniques.

This paper addresses the problem of learning binary codes that preserve similarity for an efficient search for similarity in large-scale image collections. We have formulated this problem in terms of zero-rotation data centering to minimize quantization error by mapping these data to the vertices of a zero-center binary hypercube and proposing a simple and efficient alternative minimization algorithm to perform this operation[1].

We show in this paper the two crucial difficulties recorded above can be moderated by together investigating cross-see learning and the utilization of snap information. The previous intends to make an inactive subspace with the capacity to think about data from unique perspectives (ie printed and visual perspectives), while the last investigates get to information broadly accessible and uninhibitedly available for comprehension of the question[2].

In this paper, we study HFL in the context of multimodal data for the search for inter-vision similarities. We present a new multimode HFL method, called Hashing Local Multimodal Parametric (PLMH), which learns a set of hash functions to adapt locally to the data structure of each mode[3].

In this paper, we investigated the problems of learning hash functions in the context of multimodal data for the search for similarity between cross-views. We present a new hash method, which refers to the collective matrix Factorization Hashing (CMFH)[4].

This document addresses the problem of large-scale image research. Three restrictions must be taken into account: search accuracy, efficiency and memory usage. First we present and evaluate different ways to add local image descriptors into a vector and demonstrate that Fisher's kernel performs better than the reference visual bag approach for any given vector dimension [5].

In this paper, we propose a new LASH (Lasemantic Sparse Hashing) algorithm to perform a search for similarity between modes using Sparse Coding and Matrix Factorization. In particular, LSSH uses Sparse Coding to acquire the most important image structures and Matrix Factorization to learn the latent concepts of the text.[6].

In DCDH, the paired dictionary for each mode is acquired with secondary information (for example, categories). Consequently, coupled dictionaries not only preserve the intra-similarity and interconnection between multimode data, but also contain dictionary atoms that are semantically discriminating (that is, data in the same category are reconstructed from atoms in the similar dictionary) [7].

In this paper, we propose a new method of cross-media recovery based on short and long-term relevance feedback. Our method focuses mainly on two typical types of multimedia data, ie image and audio. Firstly, we created a multimodal representation through a statistical statistical correlation between the image arrays and audio entities, and we defined the metric of the distance between the means for the measurement of similarity; therefore we propose an optimization strategy based on relevant feedback, which combines the results of short-term learning and long-term accumulated knowledge in the objective function [8].

We present a model that generates descriptions of the natural language of images and their regions. Our approach takes advantage of image data sets and their sentence descriptions to know the intermodal correspondences between language and visual data. Our alignment model is based on a new combination of convolutional neural networks on image regions, bidirectional recurrent neural networks on sentences and a structured goal that aligns the two modalities through a multimodal inlay [9].

In this article we present a new multimedia recovery paradigm to innovate large-scale research of heterogeneous multimedia data. It is able to return results from different types of media from heterogeneous data sources, for example by using a query image to retrieve relevant text documents or images from different data sources [10].

III. Proposed Methodology

We propose a new hashing technique, called semantic cross-media hashing (SCMH), to perform almost duplicate detection and cross-media recovery activities. We propose to use a set of words embeddings skip gram to represent text information. The Fisher kernel structure is incorporated to represent textual and visual information with fixed-length vectors. To map Fisher vectors in different ways, a network of deep beliefs is proposed to carry out the task. We evaluated the proposed SCMH method in two commonly used data sets. SCMH achieves better results than more advanced methods with different lengths of hash code and displays query results in order of classification.

A. Architecture

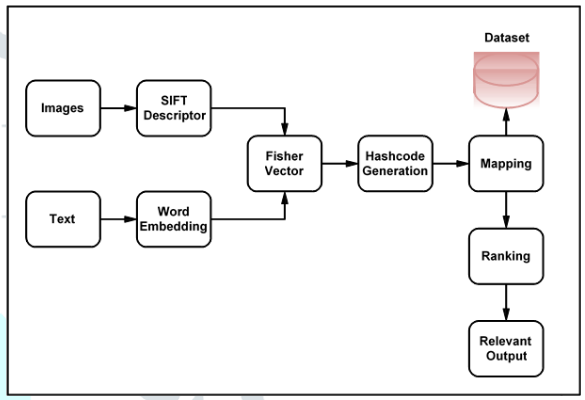

Fig. 1. System Architecture

B. Dataset

Flickr: The MIR Flickr data set, which comprises of One million images with their marks allocated by the user, It was gathered by Flickr Of the considerable number of pictures, 25,000. The pictures are annotated for 24 ideas, including the object categories i.e. tag (eg birds, people) and scene categories (for example night sky). An increasingly thorough record was made on 14 ideas where a subset of positive pictures was chosen Only if the idea is important in the picture. subsequently, This prompts a sum of 38 ideas for this informational collection. following Previous works, each picture can have a place with at least one ideas. The sets of content pictures are viewed as Similar on the off chance that they share a similar idea.

C. Algorithms:

1. Feature Descriptor

SIFT Descriptor is used for representing Images, we use SIFT detector to extract image key points. SIFT descriptor is used to calculate descriptors of the extracted key points and a variable size set of points in SIFT descriptor space represents each image.

1. The procedure to search in a repository R with query image Q.
2. The input for this operation on the user side is IDR, Q, repository key rkR, and parameter k (the number of most similar results to be returned).
3. User U starts by generating Qs searching trapdoor CQ, through IES-CBIR.
4. Then sends it to the cloud server, along with k and IDR, as parameters for the Search remote invocation.
5. The cloud starts by extracting CQs feature-vector, stems it against CBR to determine its visual words vwCQ, and accesses IdxR with them to retrieve the respective posting lists PLvw.
6. Then, for each image referenced in each of the posting lists retrieved, the cloud calculates its scaled tf-idf score and adds it to the set of results for the query. In this set, scores for the same image but different visual word are summed.
7. Finally, the cloud sorts this set by descending score and returns the results to user.

2. Word Embedding

Skip-gram algorithm is used for word embedding. After Skip-descriptor, a variable size set of points in the embeddings space represents the text.

3. Hashcode Generation

MD5 algorithm is used hash function producing a 128-bit hash value.

The MD5 message-digest algorithm is a widely used cryptographic hash function producing a 128-bit (16-byte) hash value, typically expressed in text format as a 32 digit hexadecimal number. MD5 has been utilized in a wide variety of cryptographic applications, and is also commonly used to verify data integrity.

Steps:

A message digest algorithm is a hash function that takes a bit sequence of any length and produces a bit sequence of a fixed small length.

The output of a message digest is considered as a digital signature of the input data.

MD5 is a message digest algorithm producing 128 bits of data.

It uses constants derived to trigonometric Sine function.

It loops through the original message in blocks of 512 bits, with 4 rounds of operations for each block, and 16 operations in each round.

Most modern programming languages provides MD5 algorithm as built-in functions.

D. Hardware and Software Requirements

Hardware Requirements:

1. Processor - Pentium III
2. RAM - 2 GB(min)
3. Hard Disk - 20 GB
4. Key Board - Standard Windows Keyboard
5. Mouse - Two or Three Button Mouse
6. Monitor - SVGA

Software Requirements:

1. Operating System - Windows
2. Application Server - Apache Tomcat
3. Coding Language - Java 1.8
4. Scripts - JavaScript.
5. Server side Script - Java Server Pages.
6. Database - My SQL 5.0
7. IDE - Eclipse

E. Mathematical Model

Tag and image X can be categorized by the gradient vector using the following function:

$$tt_{\theta}^{\gamma} = \nabla logP(X|\theta) = \left(\frac{\partial}{\partial\theta 1} log(P(X|\theta)) .... \frac{\partial}{\partial\theta 1} P(X|\theta)\right)$$

Where,

G_{θ}^{X}

is a vector whose dimensional is only dependent on the number of parameters, not on the number of words or key points.

The gradient describes the contribution of each individual parameters to the generative process. It can also be interpreted as how these parameter contribute to the process of generating an example. We follow the work described in for normalizing these gradients by incorporating Fisher information matrix (FIM) Fθ

$$F\theta = E(\nabla\theta logP(X|\theta)\nabla\theta logP(X|\theta)^{T})$$

Based on the specific probability density function GMM, which we used in this work, FV of X is respect to the mean m and standard deviation s of all the mixed Gaussian distributions.

Gaussian k:

$$Y_{xi}(k) = P(k|xi,\theta) = \frac{wip_k(xi|\theta)}{\Sigma_{j=1}^{N} W_j P_j(xi|\theta)}$$

IV. Result and Discussion

Experimental evaluation is done to compare the proposed system with the existing system for evaluating the performance. The simulation platform used is built using Java framework (version jdk 8) on Windows platform. The system does not require any specific hardware to run; any standard machine is capable of running the application.

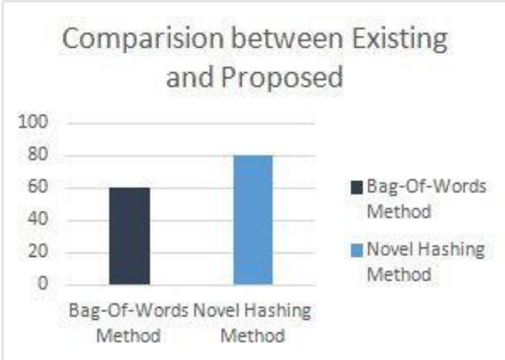

Fig. 2. Graph

Tasks	Methods	Code Length(128 bit)
Tag to Im-Age	Existing:CVH and LSSH Proposed:SCMH	Existing :0.58 Proposed : 0.60
Image to Tag	Existing:CVH and LSSH Proposed:SCMH	Existing :0.57 Proposed : 0.61

Table 1:Comparative Result

V. Conclusion

In this article, propose a new SCMH hashing method for duplicate and cross-media recovery. We are proposing to use an embedded word to represent textual information. Fisher Framework Kernel used to represent both textual and visual information with fixed-length vectors. To map Fisher vectors in different ways, a network of deep beliefs intends to perform the operation. We appreciate the proposed SCMH method in the Mriflicker dataset. In the Mriflicker dataset, SCMH on other hashing methods, which handles the best results in these data sets, are text to image and image to text tasks respectively. The experimental results demonstrate the effectiveness of the proposed method in the cross-media recovery method.

References

[1] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval, IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 12, pp. 29162929, Dec. 2013.

[2] Y. Pan, T. Yao, T. Mei, H. Li, C.-W. Ngo, and Y. Rui, Clickthrough-based cross-view learning for image search, in Proc. 37th Int.ACMSIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 717726.

[3] D. Zhai, H. Chang, Y. Zhen, X. Liu, X. Chen, and W. Gao, Parametric local multimodal hashing for cross-view similarity search, in Proc. 23rd Int. Joint Conf. Artif. Intell., 2013, pp. 27542760.

[4] G. Ding, Y. Guo, and J. Zhou, Collective matrix factorization hashing for multimodal data, in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 20832090.

[5] H. J_egou, F. Perronnin, M. Douze, J. S anchez, P. P erez, and C. Schmid, Aggregating local image descriptors into compact codes, IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 9, pp. 17041716, Sep. 2011.

[6] J. Zhou, G. Ding, and Y. Guo, Latent semantic sparse hashing for cross-modal similarity search, in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 415424.

[7] Z. Yu, F. Wu, Y. Yang, Q. Tian, J. Luo, and Y. Zhuang, Discriminative coupled dictionary hashing for fast cross-media retrieval, in Proc. 37th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2014, pp. 395404.

[8] H. Zhang, J. Yuan, X. Gao, and Z. Chen, Boosting cross-media retrieval via visual-auditory feature analysis and relevance feedback, in Proc. ACM Int. Conf. Multimedia, 2014, pp. 953956.

[9] A. Karpathy and L. Fei-Fei, Deep visual-semantic alignments for generating image descriptions, in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Boston, MA, USA, Jun. 2015, pp. 31283137.

[10] J. Song, Y. Yang, Y. Yang, Z. Huang, and H. T. Shen, Inter-media hashing for large-scale retrieval from heterogeneous data sources, in Proc. Int. Conf. Manage. Data, 2013, pp. 785796.